DOCUMENT RESUME

ED 202 420 HE 013 983

AUTHOR Wergin, Jon F.
TITLE Predicting Admissions Committee Behavior in a Medical School.
PUB DATE Apr 81
NOTE 17p.; Paper presented at the Annual Meeting of the American Educational Research Association (Los Angeles, CA, April 13-17, 1981).

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Admission Criteria; *Admissions Officers; *College Applicants; Competitive Selection; Decision Making; *Evaluation Criteria; Higher Education; In State Students; *Medical Schools; Multiple Regression Analysis; Out of State Students; *Predictor Variables; Student Characteristics
IDENTIFIERS *Medical College of Virginia

ABSTRACT

The decisions made by admissions committee members of the Medical College of Virginia were studied to determine the criteria used to arrive at value judgments and to analyze variations in predicted ratings based on these criteria. All 983 applicants to the 1980-81 entering class of the medical school who underwent file review evaluations (the total applicant pool was 2,315) were studied. The first study phase consisted of a series of stepwide multiple regression equations for each rater, using file review ratings and interview ratings as criterion variables. Predictor variables for these regressions included age, sex, minority status, presence of graduate degrees, previous applications to medical school, scores on the Medical College Admission Test, and undergraduate grade point average both in science courses and overall. Regressions were calculated both for individual committee members and the committee as a whole, and residents and nonresidents were analyzed separately. In the second study phase, each committee member's file review regression equation was applied to a random sample of 50 residents and 50 nonresidents drawn from the pool of 983 files. Results were then analyzed for the difference between the highest and lowest pairs of predicted ratings and for the placement of the actual ratings within the distribution of predicted ratings. As expected, the prediction formulas suffered from a lack of precision, due in part to the existence of certain qualitative criteria (e.g., quality of undergraduate school) not entered into the analysis. The amount of ratings variance was generally higher for individual members than for the committee as a whole. For file reviews, ratings assigned to resident applicants tended to be more predictable on the basis of admissions characteristics than those assigned to nonresidents. Additional findings and policy implications are considered. (SW)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED202420

Predicting Admissions Committee Behavior in a Medical School

Jon F. Wergin

Virginia Commonwealth University

Medical College of Virginia

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Jon F. Wergin

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

Presented to the annual meeting of
the American Educational Research Association
Los Angeles, California
April 14, 1981

HE 013 943

Predicting Admissions Committee

Behavior in a Medical School

The medical education literature is filled with studies attempting to predict performance in medical school based on known characteristics of the applicant at the time of admission. Attention lately has focused on the New MCAT as a predictor of later academic performance (e.g., Cullen, et al., 1980). A major problem with these studies, of course, is that such predictions are necessarily based on those students who have already been accepted to medical school. Research concerned with studying the process by which these students are chosen has been less frequent, although, it may be argued, the issues involved are no less important. How consistently - and thus how fairly - do medical school admissions committees make their selections? More to the point, to what extent are applicants' chances for admission based on "luck", or worse, systematic evaluator bias that is essentially unrelated to personal or academic qualifications?

Interest in this topic has increased somewhat in recent years, with most of the attention focusing on the admissions interview. Markert and Shores (1980), for example, analyzed the "difficulty" and "consistency" of interviewer/raters, and estimated that almost one-third of the candidates were unlucky enough to have been interviewed by pairs of raters with "unsatisfactory" difficulty/consistency characteristics. Milstein, et al. (1980) found that three blocks of variables ("interviewer rating tendencies", "applicant characteristics", and "applicant-interviewer dissimilarities") significantly added to the prediction of interview ratings. These recent studies strongly suggest that ratings of applicants reflect much more than applicant characteristics and random error: they also reflect systematic predispositions on the part of the evaluators, characteristics over which the applicant has little control.

But committee decisions are based on more than an interview. They are also based on a review of an applicant's "paper" credentials. Milstein, et al. (1976) attempted to predict "first-stage" (screening) decisions by selecting certain characteristics from applicants' folders and factor analyzing them. The resulting factors (MCAT scores, age, grades, "hobbies/activites", and jobs) were then used to statistically predict "accept/reject" decisions. While the accuracy of prediction was reasonably high (77%), the only significant predictors were MCAT scores and undergraduate grade point average. None of the other factors made a consistent difference.

The purpose of the current study was to explore some of the uncomfortable implications raised by such findings, specifically:

1. To what extent do individual admissions committee members differ from one another in the relative emphasis given to certain applicant characteristics, and how consistently do they apply these criteria?

2. If committee members do in fact differ, what is the possible range of individual reactions to the same applicant?

The Admissions Process at MCV School of Medicine

The admissions procedure at MCV is fairly typical of that employed at most American medical schools. Applicants are considered for a total of 168 available spaces in the first year class. Nonresident applicants are screened on the basis of undergraduate grade point average and MCAT scores. Files of those nonresidents who "pass" the screen, along with the files of virtually all Virginia residents, are then independently evaluated by two of the sixteen members of the Admissions Committee. File reviewers assign overall evaluation ratings on a nine-point scale, ranging from "A" (highly recommended) to "I" (not recommended). The two ratings assigned to each file are then brought before the full Committee, where an interview-reject decision is made. Successful candidates are independently interviewed by two Committee members, usually not the same individuals who reviewed the candidate's file, and are again rated on the nine-point "A" through "I" scale. The resultant pair of interview ratings, along with the earlier-assigned file review ratings, are then brought before the full Committee, which votes to "accept", "hold for comparison later", or "reject". In order to insure representation of multiple points of view, neither file reviewers nor interviewers are bound by a formal set of Committee-wide criteria. During the course of Committee deliberations common criteria do tend to emerge, however. As both file reviews and interviews are undertaken voluntarily, not assigned, no conscious attempt is made to "match" any particular applicant with any particular reviewer. Selection of candidates for review by Committee members is, therefore, virtually random.

The Admissions Committee consists of 12 faculty members from both clinical and basic science departments at MCV, plus 4 students. Faculty members are appointed for staggered three-year terms; students are appointed for one year only. Thus, approximately 50% of the total Committee membership is new every year.

Method

Subjects included all 983 applicants to the 1980-81 entering class of the MCV School of Medicine who underwent file review evaluations (total applicant pool = 2,315). Just under two-thirds of these applicants (n = 623) were also interviewed.

Precedure. The study proceeded in two phases. First, a series of stepwise multiple regression equations was computed for each rater, using file review ratings and interview ratings as criterion variables. Predictor variables for these regressions included the following: age, sex (coded as 0 for males, 1 for females), minority status (none claimed = 0, ethnic or financial = 1, both = 2), presence of graduate degrees

(none = 0, master's = 1, doctorate = 2), previous application(s) to medical school (0 = no, 1 = yes), MCAT-"knowledge" scores (mean of Biology, Chemistry, Physics, Science Problems scores), MCAT-"Skill" scores (mean of Reading and Quantitative scores), and undergraduate grade point average (both in "science" courses and overall). Also included as a predictor variable was a composite "letters of recommendation" score, which had previously been coded by the admissions officer and converted to a nine-point rating scale. Regressions were calculated both for individual Committee members and the Committee as a whole, and residents (Virginians) and nonresidents (non-Virginians) were analyzed separately.

In the second phase of the study each Committee member's file review regression equation was applied to a random sample of 50 residents and 50 nonresidents drawn from the pool of 983 files. A total of 16 predicted file review ratings (one for each member of the Committee) was thus generated for each of the 100 applicants sampled. Results were then analyzed in two ways:

1. For the difference between the highest and lowest pairs of predicted ratings (to determine the range of probable reactions); and

2. For the placement of the actual ratings within the distribution of predicted ratings.

Results

Phase I. First, to check for the presence of sampling bias, summary statistics were calculated for each of the predictor variables, stratified by Committee member. No significant differences were found in any of these variables across either file reviewers or interviewers, thus implying that Committee members were individually reviewing roughly equivalent samples of applicants. Significant differences were noted, however, in file review and interview ratings: 10 of 16 Committee members had mean file review ratings, and 5 of the 16 members had mean interview ratings, more than two standard errors away from the Committee mean. The reputation that some Committee members had earned for being "easy" or "hard" evaluators was therefore not altogether undeserved.

As mean ratings differed significantly from one Committee member to another, so, apparently, did the criteria used to generate these ratings differ. Individual regressions for Committee members' file reviews are given in Table 1; individual regressions for interviews are given in Table 2.

These tables reveal a number of interesting findings, both anticipated and serendipitous:

1. Perhaps most striking about these data are the wide differences observed among Committee members, both in the amount of ratings variance explained (R^2) and in the nature of the predictor

variables. File review predictability ranged from 23% to 81% for resident applications, and 13% to 53% for nonresidents. Clearly some Committee members used quantitative admission characteristics far more consistently in their evaluations than did other members; what is less clear is whether reviewers with relatively low values of R^2 were necessarily less "reliable": they may simply have been paying more attention to other, "softer" indices (such as "quality of expression", "degree of experience", and "perceived commitment"). Whether these other variables could possibly account for so much "unexplained" variance (more than 75% in one case) is open to dispute.

2. The MCAT-knowledge scores turned out to be significant predictors of file ratings in all but two regressions for residents, and all but six regressions for nonresidents. Moreover, in about 50% of the cases MCAT-knowledge was selected as the single best predictor. The next most often selected predictor was the composite "recommendation" score. The other predictor variables were spread fairly evenly across Committee members, significant in some cases and nonsignificant in others.[1]

3. The amount of ratings variance accounted for by the regression (R^2) is generally higher for individual members than for the Committee as a whole. These findings are not surprising. Individual judgments should, as a rule, be more consistent than combinations of individual judgments. It should be noted, however, that for both residents and nonresidents approximately 25% of the Committee members were actually less predictable than all sixteen members analyzed as a group.

4. For file reviews, ratings assigned to resident applicants tend to be more "predictable" on the basis of identified admissions characteristics than those assigned to nonresidents. This also is not unexpected: since nonresidents have already been preselected to some degree based on their MCAT scores and grade point averages, the variance in these two variables - and thus their predictive power - is lower. Presumably, Committee members give greater emphasis to other characteristics (e.g., essays, experience, etc.) of nonresidents' files - although there is no way of inferring that directly from the data. An equally plausible hypothesis is that nonresidents are simply rated less consistently.

---

[1] Absence of a particular variable as a significant "predictor" in a stepwise regression equation does not necessarily indicate lack of relationship with the criterion variable. For example, the MCAT-skill variable was a significant predictor in only one of the individual regressions for resident applicants, even though it displayed a substantial correlation with file reviews by itself ($r = .27$). But since MCAT-skill is highly correlated with MCAT-knowledge, and MCAT-knowledge itself was a slightly better predictor ($r = .37$), MCAT-skill typically had little incremental predictive power.

5. In the main, prediction of interview ratings (Table 2) paralleled file reviews. Values of R^2 were, however, uniformly lower, ranging from 6% to 67% for residents, and 2% to 44% for nonresidents. Since the primary function of the interview is to gain a personal impression of the applicant as a future physician, an observed weaker relationship between quantitative variables and interview ratings comes as no surprise. One might argue, in fact, that an R^2 of 67% is actually high given the purported emphasis on personal qualities in the interview setting.

Phase II. As noted previously, a second series of analyses was undertaken to determine if, in fact, observed differences among Admissions Committee members would likely result in significantly different ratings of the same applicant: to assess, in other words, the extent to which an applicant's chances for admission are determined by the "luck of the draw".

The range of predicted file review ratings for the sample of 100 applicants was startling. First, in all 100 cases, the two highest predicted ratings would have virtually guaranteed an invitation to interview. But for the nonresident sample, the two lowest predicted ratings would have virtually guaranteed rejection, in 46 of the 50 cases. The range of prediction was somewhat smaller in the resident sample: here, the two lowest predicted ratings would have virtually guaranteed rejection in 23 of the 50 cases. Thus, for the entire sample of 100 files, more than two-thirds (69%) could have conceivably led to either "interview" or "reject" decisions, depending on the identity of the two reviewers. But these findings, while disturbing and thought-provoking, are misleading. Of the 46 nonresidents facing contradictory predictions, for example, one of the two lowest predicted ratings came from the same Committee member 66% of the time. This individual had developed a reputation on the Committee for being a singularly difficult grader (a reputation borne out, incidentally, by the data), and thus the full Committee tended to adjust his ratings upwards (e.g., "A rating of F from Dr. X is like a D from the rest of us"). Further, this person reviewed a relatively small number of files (23), and was relatively unpredictable in his assigned ratings (R^2 = .28). In truth, therefore, an analysis of extremes may imply a far greater range of probable reactions than actually existed.

The second analysis - an inspection of actual file review ratings within the distribution of predicted ratings - proved to be more revealing. The analysis procedure was as follows. First, a median predicted rating was calculated for each of the 100 sample applicants. Next, a decision rule was formulated, based on past behavior of the Committee, such that a median rating greater than 4.0 (equivalent to two file ratings of "E") was high enough to warrant a decision to interview a resident applicant, and a median rating greater than 4.5 (equivalent to file ratings of "D" and "E") was high enough to warrant an interview decision for a nonresident applicant. Predictions of "interview" or "reject" were then made for each sample applicant using this criterion, and the results were compared to the Committee's actual decision (based, of course, on the actual file reviews). Results of this analysis are given in Table 3.

As these data indicate, in just over two-thirds of both the resident and nonresident samples (34/50), the predicted decision matched the actual decision. Errors of prediction were biased in the resident sample toward the "predict interview/ actual reject" category, while errors of prediction for nonresidents were evenly split. The ultimate fate of those interviewed was then examined, to see how closely the eventual accept/reject decision matched the file review prediction. These findings are displayed in Figure 1.

As the data from this figure show, 64% of the applicant sample was interviewed. About one in five interviews had a predicted file review rating of "reject", and most of this latter group (8% of the total file review sample) was in fact eventually rejected. Using this figure plus or minus two standard errors (s.e. = .03), one might conclude that between 2% and 14% of the total pool should never have been interviewed at all. Estimating the percentage of applicants who were possible victims of a negative five reviewer bias is much more difficult. Twenty-two percent of the sample were rejected at the file review stage, even though predicted ratings by other Committee members suggested an "interview" decision. How many of these applicants would likely have earned higher ratings from a different pair of raters? How many, in contrast, would have been rejected anyway, for reasons related to criteria not included in the study (e.g., low perceived motivation, poor self expression, undesirable geographic location)? Given the substantial "unexplained" variation in the individual prediction equations, a definitive answer - even an educated guess - is impossible. Some of these applicants would have, no doubt, been accepted had they been granted an interview; but the size of this group is unknown.

Conclusions

So what do the data reported in this study tell us about the differential effects of admissions committee members' individual judgments? Much - and little. Certainly, the behavior of individual Committee members in evaluating applicants varies markedly, in at least three ways: the average ratings they assign, the relative emphasis paid to known applicant characteristics, and the consistency with which those characteristics are used as decision criteria. Further, given these variations, a wide range of individual reactions to the same applicant is not only possible but likely. Since an applicant's file is evaluated by only two of the sixteen Committee members, an obvious question of fairness in selection arises. Here is where the data are less clear - and less useful. A number of mitigating circumstances weaken the strength of any "fairness" argument, thus leaving the issue unresolved:

1. As expected, the prediction formulas suffer from a lack of precision, due at least in part to the existence of certain qualitative criteria (e.g., "quality of undergraduate school") not entered into the analysis. Any prediction "error" is

therefore traceable in part to the evaluator's attention to these characteristics. The methodology employed in this study has obvious limitations, relating mostly to its reliance on various kinds of statistical analyses. Precision would be strengthened by conducting a followup clinical analysis of cases where predicted decisions did not match actual decisions. Evidence of commonality could be sought, for example, from the files of those applicants falling into the "predicted interview but rejected" category, and compared with similar evidence collected from the files of "predicted rejection but interviewed" applicants. A comparison of these commonalities could conceivably result in a more exact determination of the relative strength of qualitative criteria.

2. Predicted ratings by file reviewers tended to cluster around the median, thus decreasing the probability of extremely high or low ratings by Committee "outliers", who tended to review fewer files anyway.

3. Third, and perhaps most important, is the circumstance surrounding the functioning of the Admissions Committee as a group. Throughout a period of approximately eight months, the Committee as a whole considered the qualifications of nearly one thousand applicants competing for 168 available spaces. There is no doubt that throughout the course of their deliberations decisions were made which, due to unfairly biased ratings or just the press of time, must be considered capricious. But an opposing force must also be considered. As noted earlier, Committee members tend to consider the source of assigned ratings, probe for apparent discrepancies, and adjust their decisions accordingly. An overt attempt is made, in other words, to balance multiple values and points of view. How successful is this attempt? The upper bound of decision "error" based on file review ratings, according to the current data, is 30% (the total of "false positive" and "false negative" predictors) and given the mitigating circumstances described above, is almost certainly lower than that. Whether an error of this magnitude is acceptable requires a value judgment best left to the reader.

References

Cullen, T.J., Dohner, C.W., Peckham, P.D., Samson, W.E., and Schwarz, M.R.: Predicting first-quarter test scores from the New Medical College Admission Test. Journal of Medical Education, 55, 393-398, May, 1980.

Markert, R.J., and Shores, J.H.: Assuring fairness in the medical school admission interview through analysis of rater difficulty and consistency. Presented to the annual meeting of the American Educational Research Association, Boston, MA, April, 1980.

Milstein, R.M., Burrow, G.N., Wilkinson, L., and Kessen, W.: Prediction of screening decisions in a medical school admission process. Journal of Medical Education, 51, 626-633, August, 1976.

Milstein, R.M., Burrow, G.N., Wilkinson, L. and Kessen, W.: Prediction of interview ratings in a medical school admission process. Journal of Medical Education, 55, 451-453, May, 1980.

Table 1

Individual Stepwise Regressions: file ratings on admission variables, by applicant resident status

Entering Class of 1980-1981

Residents/Nonresidents

Committee Member+	PREDICTOR										R^2
	Age	Sex	Minority	Graduate Degree	Previous Application	MCAT Knowledge	MCAT Skill	GPA Overall	GPA Science	Recommendations	
A	*/					*/		/*		*/	.61/.36
B						*/				*/*	.59/.43
C	*/	*/	*/			*/*	/*	/*	*/	*/*	.58/.40
D		*/*				*/*	/*	*/		/*	.81/.52
E		*/	*/*			*/*		*/		*/*	.23/.13
F					*/	*/*				*/*	.46/.44
G	*/	*/*		*/*		*/*	/*		*/*	*/*	.51/.35
H	/*			/*	*/	*/		*/	/*		.55/.44
I			/*			*/*		*/*		/*	.46/.51
J			*/*			*/	/*		*/*	*/	.44/.53
K	*/		*/*			*/*		/*	*/	*/*	.53/.30
L	/*					*/	/*		*/		.44/.16
M							*/		/*		.31/.28
N						*/*		*/*		*/*	.38/.26
O				/*		/*			*/		.34/.36
Whole Committee	*/	*/*	*/*	*/		*/*	*/*	*/*	*/	*/*	.39/.27

* Significant predictor at $P < .05$

\+ Number of files reviewed for one committee member was too small to allow meaningful estimate.

Table 2

Individual Stepwise Regressions: interview ratings on admission variables, by applicant resident status

Entering Class of 1980-1981

Residents/Nonresidents

Committee Member	PREDICTOR										R^2
	Age	Sex	Minority	Graduate Degree	Previous Application	MCAT Knowledge	MCAT Skill	GPA Overall	GPA Science	Recommendations	
A							*/			/*	.21/.18
B				/*		*/					.10/.10
C		/*				*/		*/		*/	.44/.14
D			*/	*/	*/	*/*	*/	*/	/*		.59/.32
E				*/	*/	/*				*/	.24/.02
F					*/			*/	/*		.19/.08
G						/*		*/			.37/.03
H	*/			*/		*/	/*				.41/.17
I						*/*		*/		*/	.60/.22
J	*/	/*	*/*	*/		*/		/*	*/	*/	.50/.44
K	*/	*/*						*/			.21/.20
L				*/		/*	*/	*/	*/		.49/.17
M								*/		/*	.06/.11
N								*/	/*	*/	.30/.10
O	*/			/*			/*				.10/.41
P		*/	/*		*/	*/		*/			.67/.14
Whole Committee	*/		*/	*/	*/*	*/*	*/	*/	/*	*/	.23/.06

* Significant predictor at $P < .05$

TABLE 3

PREDICTED VERSUS ACTUAL DECISIONS

BY APPLICANT RESIDENCE STATUS

Actual

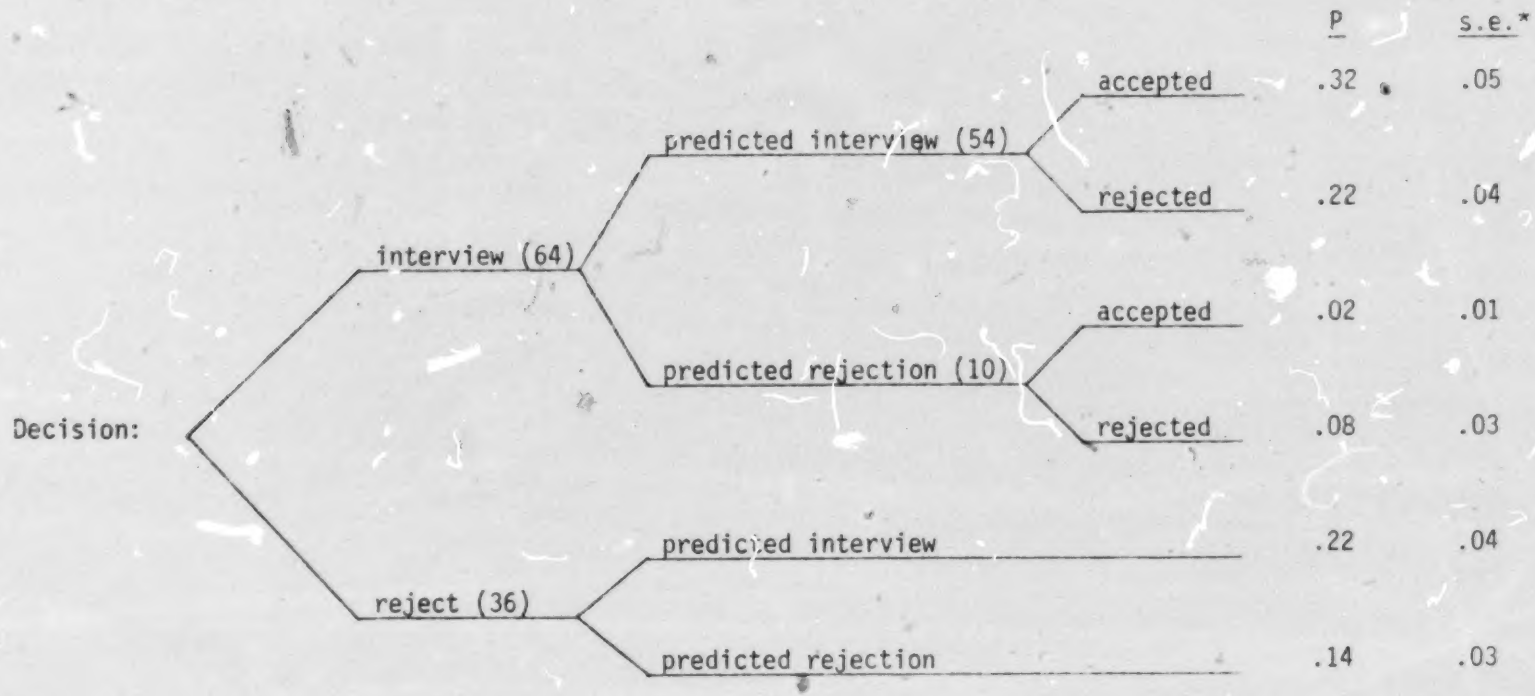
Predicted		Interview	Reject	Total	
	Interview	30	14	44	RESIDENTS
	Reject	2	4	6	
	Total	32	18	50	

Predicted		Interview	Reject	Total	
	Interview	24	8	32	NON-RESIDENTS
	Reject	8	10	18	
	Total	32	18	50	

FIGURE 1

ANALYSIS OF ADMIT/REJECT DECISIONS

(n = 100)



* Standard error of the proportion.